Deepfake technikák terjedése: harc az újgenerációs kibercsalások ellen

A mesterséges intelligencia eszközeinek elterjedésével ma már minden eddiginél könnyebb megszemélyesíteni valakit és megtéveszteni másokat, pénzt, adatot kicsalni a célpontoktól. A korábban kifejlesztett módszerek hatékonyságát megsokszorozza az AI, melyet a vezető intézmények irányelvek és technológiák kombinációjával próbálnak kezelni – áll a Deloitte legfrissebb Tech Trends elemzésében.

Napjainkra a mesterséges intelligencia és deepfake elérte azt a minőséget, hogy bárkit úgy lehessen beállítani, mintha valamit mondott vagy tett volna. A generatív AI gyors fejlődésének köszönhetően a mesterségesen generált tartalom elérte azt a pontot, hogy nehéz megkülönböztetni, mi a valós és mi nem.

> Az AI eszközök elterjedésével a kiberbűnözők minden eddiginél könnyebben adják ki magukat másnak és tévesztik meg áldozataikat. Sokan a hang- és arcfelismeréses hozzáférés-ellenőrzés megkerülésére, valamint adathalász-kísérletekhez deepfake-et használnak. Kedvelt célpontjuknak számít a hatalmas mennyiségű adatot igénylő AI-alkalmazások, a biztonsági kockázatok ezért megsokszorozódnak minden egyes új tartalomgeneráló eszközzel, amely az internetre kerül – mondta **Szöllősi Zoltán**, a Deloitte Magyarország kockázatkezelési tanácsadás üzletágának partnere.

A pszichológiai manipulációt alkalmazó bűnözők fő módszere, hogy meggyőzzék a célpontot, aki hozzáférést enged a rendszerekhez. Ez a stratégia ugyan nagyon hatékony, de sok személyes interakciót igényel a csaló és az áldozat között. A mesterségesen generált tartalom lehetővé teszi a támadók számára, hogy sokkal kisebb időbefektetéssel hozzák létre ezt a személyes kapcsolatot: a mesterségesen generált tartalmak mostanában a cégek biztonsági réseit veszik célba, megbízható forrásoknak kiadva magukat.

A Deloitte felmérésében a megkérdezettek többsége úgy véli, képesek megkülönböztetni az AI és a valós emberek által létrehozott tartalmat, és mindössze 20% kételkedik ebben. Ez abból adódik, hogy az emberek arra számítanak, hogy a mesterségesen generált tartalom valamilyen módon mesterségesnek néz ki vagy annak hangzik. Azonban a valóságban kevesen tudnak megbízhatóan különbséget tenni a kettő között, mivel az AI tartalomgenerátorokat az ember által létrehozott tartalmakon tanítják.

Négy alapvető kiberbűnözői AI-módszer

Javított adathalászat: Az adathalászat a kibertámadások leggyakoribb típusa, 2021-ben a becslések szerint 44,2 millió dollárt loptak el a módszerrel. Az adathalász-támadások általában nem az e-mailek minősége, hanem sokkal inkább a tömeges kiküldések miatt sikeresek. A legtöbb címzett képes felismerni a kísérleteket, pl. a rossz nyelvtan miatt vagy mert a feladót nem ismerik. A generatív AI-eszközökkel azonban gyorsan és egyszerűen, egyre meggyőzőbb üzeneteket lehet készíteni. Ráadásul a nyilvánosan elérhető MI-modellek minőségének javulásával a probléma csak egyre súlyosabb lesz.

Deepfake: A deepfake tartalmak évek óta léteznek, de egészen a közelmúltig nem voltak elég

meggyőzőek ahhoz, hogy kiber-bűncselekményekben használják őket. Azonban egyre gyakrabban tapasztalható, hogy vállalkozások elleni támadásokat építenek rájuk. Egy brit energiaipari vállalat vezérigazgatójától például 243 000 dollárt szereztek meg csalók, akik a deepfake AI hangtechnológiát használva az anyavállalat vezetőjének adták ki magukat.

Prompt injektálás: A hackerek a virtuális asszisztensekkel rendelkező böngészőket és e-mail klienseket is használják. Ebben az esetben olyan rosszindulatú promptokat adnak ki, melyek arra utasítják a virtuális asszisztenst, hogy adatokat, banki, egészségügyi információkat továbbítson - automatikusan, az áldozat tudta nélkül.

Hamis információterjesztés: A cégek elleni közösségi médiakampány nem újdonság, AI-eszközökkel azonban gyorsan hatalmas mennyiségű negatív tartalom hozható létre, mellyel a kiberbűnözők célba vehetnek vállalkozásokat. Míg korábban a támadóknak személyesen kellett ezeket a tartalmakat elkészíteniük, jelenleg MI-eszközökkel sokkal nagyobb mértékű félretájékoztatást lehet produkálni.

> A generatív AI széleskörű elérhetősége és a tartalomgeneráló modellek fejlődésének üteme valószínűleg nagymértékben felerősíti a már korábban is észlelt problémákat. Kevés vagy nulla költséggel és gyakorlatilag technikai készségek nélkül bárki képes lesz olyan meggyőző anyagot létrehozni, amellyel pénzt vagy adatot tulajdoníthat el. Jó hír, hogy ugyanazok az eszközök, amelyeket a bűnözők az intézmények kifosztására használnak, felhasználhatók a támadások azonosítására, előrejelzésére és megelőzésére is. Éppen ezért az intézmények a káros tartalmak azonosítására és az alkalmazottak kockázatokkal kapcsolatos oktatására kidolgozott irányelvek és technológiák kombinációjával reagálnak – mondta **Szöllősi Zoltán**.

Hogyan védekezzenek a vállalatok az új fenyegetés ellen?

Amikor a generatív AI-eszközök nyilvánossá váltak, a hackerek előnyt élveztek, mivel ezek a modellek a legnagyobb technológiai cégek legerősebb hardvereihez és legnagyobb méretű adathalmazaihoz fértek hozzá. A felismerő rendszerek első generációjának nem volt megfelelő nagyságú adathalmaza a szintetikus tartalmak azonosítására ám ez mára megváltozott. Bizonyos vállalatok egy petabájt méretű szöveg-, kép- és hangadatbázison tanítják szintetikus médiafelismerő platformjukat, amelynek egy része mesterségesen generált. Egy ekkora adathalmazon alapuló vizsgálat során már apró jelek is mutathatják, hogy valamit AI-eszközzel hoztak létre.

Más eszközök is vannak már MI által generált tartalmak felismerésére: az Intel nemrég mutatta be deepfake-felismerőjét, amely az adatokon túlmenően a videókon szereplő emberek arcán való véráramlást figyelve elemzi a videókat. Az ember szívének pumpálásával az erek színe enyhén megváltozik, amelyet az AI-modellek számára nagyon nehéz jelenleg utánozni.

Felkészül: a kvantumszámítógép

A kvantuminformatika még néhány évnyire van attól, hogy széleskörben elérhetővé váljon, a folyamat azonban gyorsul és könnyen lehet, hogy ez lesz a hackerek és a vállalatok harcának következő eszköze. A technológia egyik legígéretesebb felhasználási területe a kvantum gépi tanulás, amely révén kevesebb rendelkezésre álló adatból is lehet előrejelző modelleket létrehozni, és sokkal összetettebb modellek kidolgozását teszi lehetővé, mint ami ma a legfejlettebb grafikus feldolgozóegységek hardverének segítségével lehetséges.

A kiberbiztonságukat javítani kívánó vállalatok számára a kvantum gépi tanulás a szintetikusmédia-

szűrőket javíthatná. Ahelyett, hogy több milliárd adatpontra lenne szükségük ahhoz, hogy megtanulják felismerni a mesterségesen létrehozott médiát, a speciális szűrők már néhány példa megvizsgálása után megtanulhatják felismerni a hamisítványokat.

Tech Trends 2024 tanulmány letöltése

Sajtókapcsolat:

- Szöllősi Zoltán, partner, információbiztonsági szakértő
- Deloitte Magyarország
- zszollosi@deloittece.com

Eredeti tartalom: Deloitte Magyarország

Továbbította: Helló Sajtó! Üzleti Sajtószolgálat

Ez a sajtóközlemény a következő linken érhető el:
https://hellosajto.hu/11572/deepfake-technikak-terjedese-harc-az-ujgeneracios-kibercsalasok-ellen/